第4章

画像圧縮処理の中身と画質への影響

―― 開発システムに合った画像エンコーダ選択の手引き

清 恭二郎

ディジタル画像を圧縮・伝送・復号化するシステムにおいて，画像の質を決定づけるのは，テレビに近い受信側ではなく，むしろ画像を圧縮するエンコーダ側にあると言われる．ここでは，その理由と，画像エンコーダの選択のヒントを述べる．

（編集部）

例えばビデオ・カメラを想定した場合，イメージ・センサからの出力は，大きく以下の三つのモジュールを経由してビット・ストリームになります．

- A-Dコンバータなどのアナログ部
- 補間処理，色変換，ガンマ補正などを行うディジタル・フロントエンド
- MPEG，H.264など，圧縮のためのディジタル・バックエンド

ここではおもに画像の圧縮処理に注目し，「アルゴリズムに起因する画質[注1]」を決定づけるカギは何であるのか，また，それを最適化するためにはどういうくふうが必要なのか，を掘り下げていきます．

1．動き予測と符号量制御が圧縮処理の肝

図1はMPEGなどの圧縮処理のブロック構成を簡単に示したものです．圧縮処理は，基本的にDCT（離散コサイン変換）やDHTといった直交変換を利用した「画面内テクスチャ符号化」と，前後のフレームから動きを検出・補償し

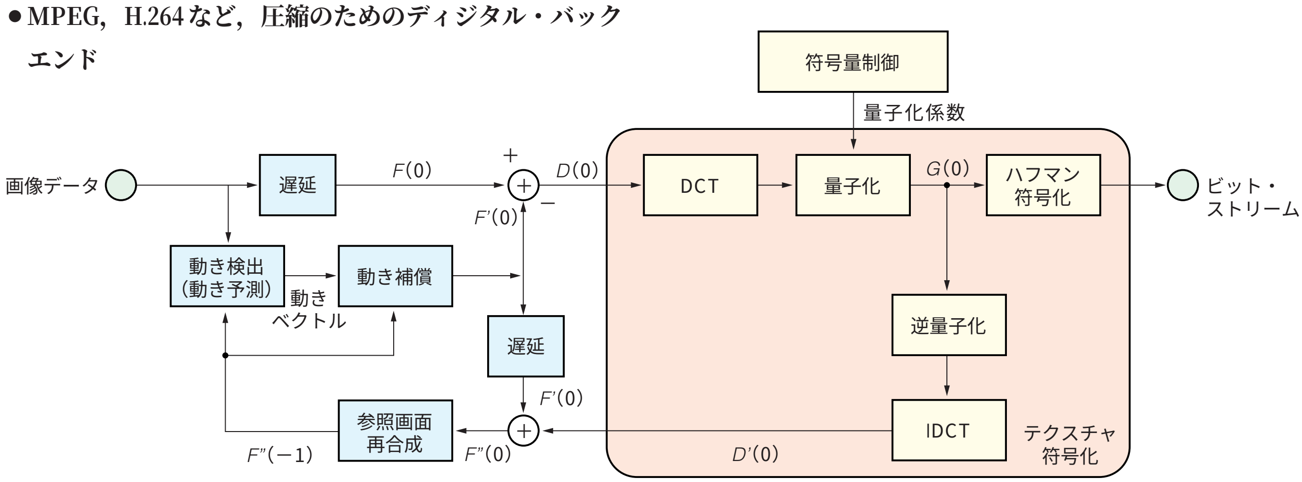


図1　圧縮処理のブロック構成
画質を左右する重要なモジュールは，「符号量制御」と「動き予測」の二つだけ

注1：ここでいう画質とは，「圧縮処理を施していない画像」と「符号化し，さらに復号したあとの画像」の見た目の違いを表したもの．圧縮処理を施していない画像に近いほど，質が高いとする．

KeyWord　動き予測，符号量制御，画面内テクスチャ符号化，フレーム・ベース，マクロブロック・ベース，1pass，2pass，CBR，VBR，GOP，探索アルゴリズム，探索範囲，ダイヤモンド・サーチ，3ステップ・サーチ，サーチ・レンジ

たうえでテクスチャ符号化を行う「フレーム間符号化」から構成されています．

図1において，入力された画像はマクロブロックと呼ばれる16×16画素を一つの単位として，あらかじめ蓄積されていた前後のフレームからの動きを検出し，その結果に基づいて動き補償を行います．その動き補償後のデータと元画像との差信号をテクスチャ符号化するのが，フレーム間符号化と呼ばれる処理です．

ここで，画質という観点からこの処理を見た場合，画質を左右する重要なモジュールは，「符号量制御」と「動き予測」の二つだけといってよいでしょう．そのほかの構成要素は基本的に定型的な算術演算処理を行っており，画質の差は発生しません．そのため，各チップ・メーカやIPコア・ベンダによるこれらの基本性能の差は，処理速度，またはチップ面積からくるコストの差と考えられます．

一方，符号量制御と動き予測が画質に与える影響には多大なものがあり，MPEG開発技術者の腕の見せどころでもあります．A社のチップは低ビット・レートでは高画質だが，高ビット・レートではB社のほうが画質は良い，あるいはC社のIPコアは動きが早い画像でも破たんしないがゆっくりした動きの画像ではD社のほうが良さそうだ，などさまざまです．そこで，この符号量制御と動き予測の2点について，さらに詳しく説明します．

● 符号量制御のアルゴリズムは大きく三つ

符号量制御とは，直交変換したデータを量子化する際に係数を決定するための制御処理で，画像の複雑さやすでに生成された符号量を考慮して最適化していきます．量子化係数が大きくなれば生成符号量は小さくなり，ビット・レートを下げられますが，再生する際に元画像との差が大きくなり，モスキート・ノイズやブロック・ノイズと言われるディジタル・ノイズが増えてしまいます．量子化係数を小さくすれば，それだけ損失がなくなるので画質は向上しますが，生成符号量は増え，ビット・レートが上がります．符号量制御のアルゴリズムの基本は，次の3種類に分けて考えられます．

1）フレーム・ベース/マクロブロック・ベース

2）1pass/2pass

3）CBR/VBR

次に詳しく説明します．

1）フレーム・ベース/マクロブロック・ベース

MPEGやH.264では，マクロブロックと呼ばれる16×16画素を基本単位として処理を行います．符号量は，マクロブロック周期，フレーム周期，GOP（group of picture）周期で制御します．

このGOPは，十数フレームから数十フレームを一単位とした画像で，GOPの中にIフレームが一つ存在し，残りをPフレームやBフレームで構成しています．Iフレームは，PフレームやBフレームよりもたくさんの符合が必要なので，GOP周期で符合量を平滑化する制御を行います．そして，フレーム単位でI，P，Bフレームの種類に応じて符合量を割り当てます．

ここでは符号量制御ループをマクロブロック単位で行うことを，マクロブロック・ベースと呼びます．マクロブロック・ベースの場合，あるマクロブロックの動きベクトルが検出されると，直ちにそのマクロブロックの動き補償を行い，テクスチャ符号化を行います．したがって，そのピクチャ全体を把握することなく符号化していくことになります．いわば，「木を見て森を見ない」的な手法です．

一方，フレーム・ベースでは，マクロブロック単位で動き予測を行っても，そのフレーム処理が終了するまではテクスチャ符号化を行いません．一般的に動き予測回路では，予測を検出していく過程でそのフレームの複雑さを解析していくことができます．あまり変化のない単調な画像か，あるいは非常に細かいテクスチャが含まれた画像か，もしくは画面の上半分は簡単だが下半分はかなり複雑である，といった特徴を解析できます．そのため，そのフレームの予測処理が終了したとき，画像の内容を理解したうえで，その時点までに生成した符号量も考慮して最適な符号配分を決めることができます．これがフレーム・ベースです．「森」全体を見ている手法です．画質を論じる場合は，フレーム・ベースのほうがはるかに良い手法と言えるでしょう．

マクロブロック・ベースではメモリ使用量や回路など，ハードウェア資源を最小限にできることがメリットと言えます．MPEGのIPベンダでは，このタイプのアルゴリズムを用いている例がいくつか見られます．超低ビット・レートの携帯電話などでは有効でしょう．

2）1pass/2pass

1pass/2passは，フレーム・ベースをさらに進めたものと言えます．2passとは圧縮を2回に分けて行うことを意味し，1回目で圧縮したい画像全体の解析を行い，2回目で

完成したビット・ストリームにします．DVDのディスク容量に合うように，映画を圧縮する工程ではこの手法を使っています．例えば2時間分の映画をすべて解析したうえで，その全体から最適な量子化係数を決めて圧縮しているのです．つまり，「森」はもちろん「世界全体」まで見ていると言えるでしょう．これは，与えられた条件下では最高の画質が得られる，と言っても過言ではありません．

それに対し，DVDレコーダやDVDビデオ・カメラは，リアルタイムに画像圧縮を行うことを前提としており，2passは使えません．そのため，市販のDVDのディスクと同じビット・レートであったとしても，画質的に同じレベルに到達するのは困難と言えます．

3）CBR/VBR

CBRはconstant bit rate（固定転送レート）を，VBRはvariable bit rate（可変転送レート）を表します．図2に転送レートの違いを示します．CBRとは，転送レートを固定にするものですが，実際にはIフレームやPフレーム，Bフレームなど，ピクチャ・タイプの違いや画像そのものの複雑さにより，生成符号量はフレームごとに違います．ここでは設定されたビット・レートを平均レートとして制御する方法をいいます．

これに対して，画像の状況に応じて転送レートを変えていくことをVBRと呼びます．画像の複雑なシーンが続く場合，実効的な転送レートを上げられるため，画質の向上が期待できます．市販のDVDディスクでは2passVBRが使われています．一方，入力画像をリアルタイムに圧縮しなければならないDVDレコーダやDVDビデオ・カメラでは1passVBRが多く使われています．これは入力画像を数フレーム分蓄積させたあとで圧縮するものです．蓄積している間に画像解析を行い，その範囲でビット・レートを最適化するフィード・フォワード制御を行っています．

● 動き予測のよしあしは三つの要素で決まる

動き予測とは，1枚の画面を16×16画素のマクロブロックに分割して，それぞれのマクロブロックが参照画像のどの位置から移動したのかを探索し，そのマクロブロックごとに一つまたは複数の動きベクトルで表現することを言います．これには膨大な演算が必要とされ，通常はハードウェアで処理されます．このとき，この動きベクトルを求める手法には多くの種類がありますが，このアルゴリズムの違いによって画質は大きく影響を受けます．このことを以下の3点に分けて考察していきましょう．

- 探索アルゴリズム
- 探索範囲
- モード決定手法

動きベクトルは，対象画像におけるマクロブロックと参照画像における任意の位置の16×16画素データとの，画素ごとの差分絶対値の総和を取り，その最小値を探索することにより求められます．その際に参照画面内の全体を探

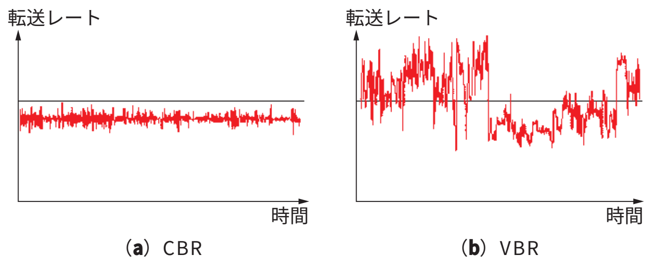


図2　転送レートの変化のしかたの例
複雑なシーンが続く場合，VBRは実効的な転送レートを上げられる

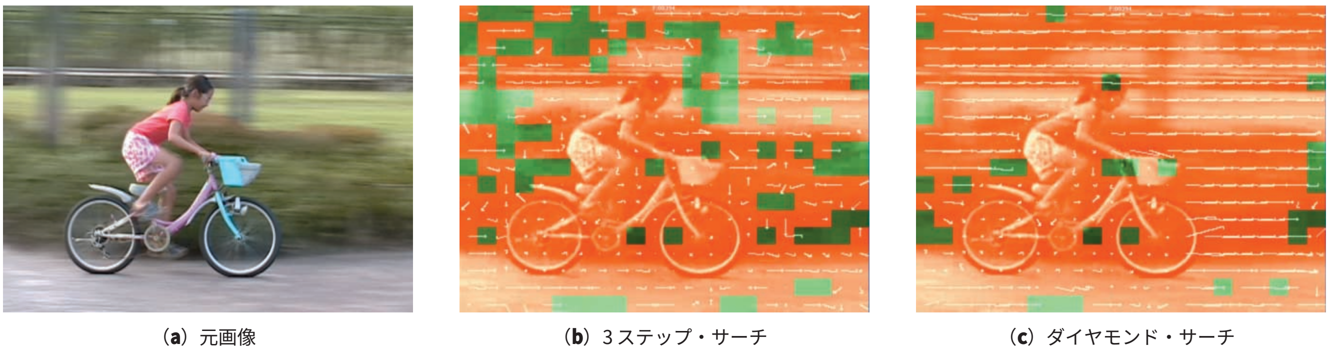


図3　アルゴリズムの差によるベクトル検出の差
白い矢印はマクロブロック単位の動きベクトルを示したもの

表1　図3のサーチにより発生した符号量

サーチの種類	ファイル・サイズ［Kビット］(Q＝8)
3ステップ・サーチ	1096
ダイヤモンド・サーチ	927

表2　同一の探索アルゴリズム（ダイヤモンド・サーチ）を用いて探索範囲を変えた時の生成符号量の差

探査範囲	±8.5	±16.5	±32.5	±64.5	−128〜＋127.5
flower	2,585,459	2,583,816	2,581,580	—	—
mobile	4,189,686	4,190,158	4,192,287	—	—
football	1,820,344	1,534,764	1,375,745	1,276,309	1,260,999
Car&Bicycle	1,951,781	1,615,276	1,390,413	1,308,817	1,310,369

表3　圧縮処理に使われるチップの品種による性能の違い

	汎用RISC CPU	DSP（専用演算あり）	専用MPEGチップ
テクスチャ符号化	1.6Gcps（クロック・サイクル/s）	180Mcps	54MHz注1
動き予測	0.8Gcps	90Mcps	54MHz注1，注2

注1：クロック周波数
注2：サーチ・レンジ±80画素

索するとよいかもしれませんが，あまりにも膨大な計算量になるため，効果的な探索アルゴリズムを検討したり探索範囲を制限したりすることにより，演算の最適化を行っています．アルゴリズムの詳細は省略します．**図3**はアルゴリズムの差によるベクトル検出の差を表しています．自転車に乗っている子供を単にパンニングして撮っただけです．**(a)**の元画像に対し，**(b)**は3ステップ・サーチ，**(c)**はダイヤモンド・サーチと呼ばれるアルゴリズムの結果です．**(b)**，**(c)**の中の白い矢印はマクロブロック単位の動きベクトルを示したものです．**(b)**は検出したベクトルが不安定ですが，**(c)**はかなり正確に動きを捕らえています．この結果，発生した符合量を比較したものが**表1**です．量子化係数を8に固定したところ，生成符号量に約10％もの差が生じました．

表2は同一の探索アルゴリズム（ダイヤモンド・サーチ）を用いて探索範囲を変えたときの生成符号量の差を示しています．これは，640×480画素におけるMPEG標準画像を利用し，100フレームを圧縮した際の生成符号量で，football や Car & Bicycle はかなり動きの激しい例，flower，mobileは動きの緩やかな例です．動きの激しいものでは，探索範囲の広いほうが有利であることがわかります．IPベンダが発表しているMPEG製品の中には，VGAクラスでありながら探索範囲が±16しかないものもあり，圧縮の目的によっては選択に注意が必要でしょう．

ところで，参照画像がかならずしも1フレーム分過去の画像とは限りません．2フレーム分かもしれませんし，未来の画像を参照することもあります．また，インターレース画像の場合にはフィールド構造で予測したほうがよいのか，フレーム構造で予測したほうがよいのか，それも判断しなければなりません．これらは動き予測回路とその結果に基づいて上位のファームウェアで決定されることが多く，各社のノウハウによるところが大きいと言えます．

2. 自分の理想に近いエンコーダを選択する

読者のみなさんの中には，これからシステムに組み込むためのMPEG圧縮用のチップを選定している方もいることでしょう．チップには大きく2種類のタイプがあります．一つはCPUや汎用DSPにマイクロコードを組み込んで実現するタイプ，もう一つは専用のハードウェアで構成された専用MPEGチップです．アプリケーションによって画質に対する要求も違うので，どちらが良いとは一概には結論づけられません．

ここでは前述のような解析結果から，必要とされる演算能力をそれぞれのタイプのチップで実現するとどうなるかを，筆者ら（ニューコアテクノロジー）の経験から算出してみました．**表3**にチップの品種による性能の違いを示します．MPEG-4シンプル・プロファイルの圧縮演算において，マクロブロック（16×16画素）を1単位として必要な演算量を算出し，それをもとに640×480画素で30フレーム/sの画像を圧縮する際に必要なパフォーマンスを予測した値です．ただし，専用チップについては実在するチップの仕様です．

まず，テクスチャ符号化だけ（動き予測を行わない）の場合は定型的な演算だけと説明しました．ところが，筆者の経験ではこの演算を32ビット汎用RISC CPUで行った場合，約1.6Gクロック・サイクル/s必要となりました．以下，専用演算機能付きDSPでは180Mクロック・サイクル/s，専用MPEGチップでは54MHzです．また，動き予測については，サーチ・レンジを±16画素として計算すると，汎用CPUでは0.8Gクロック・サイクル/s，DSPでは

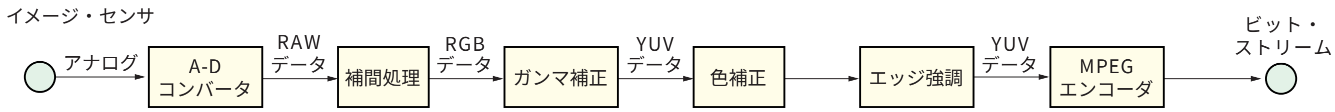


図4　画像エンコーダの圧縮の前処理
イメージ・センサから出力された信号は，A-Dコンバータを経てさまざまな画質調整を行った後に圧縮処理をする

90Mクロック・サイクル/sとなりました．

前述したように，640×480画素の場合，高画質をねらうのであればサーチ・レンジは64画素くらいに広げたいところです．この場合，単純に動き予測を行うと**表3**の値のさらに64倍の性能が必要となり，ますます実現が難しくなります．サーチ・レンジを広げる場合，演算量を減らすために階層的に行うなど，さまざまなくふうがなされていますが，いずれにしても汎用CPUまたはDSPでは実現は困難と言わざるをえないでしょう．一方，専用MPEGチップでは54MHzで±80画素を実現しています（ニューコアテクノロジーの「SiP-2290」の場合）．なお，この専用チップは，動き予測の回路とテクスチャ符号化の回路が独立に存在しているため，全体でも54MHzで640×480画素，30フレーム/sの画像をエンコードすることが可能となっています．

3．圧縮する前の画像の質も重要

今までは圧縮処理の中身と画質との関係を論じてきましたが，ここではビデオ・カメラを例に，圧縮の前処理として画質を決めるかぎは何かを説明しましょう．

本稿の冒頭でも述べたように，イメージ・センサから出力された信号は，A-Dコンバータを経てさまざまな画質調整を施したあと，圧縮処理を行います．**図4**にそのシステム・ブロック図を示します．イメージ・センサから出力された生データは，補間回路でRGBデータに変換されます．この後，色調整，ガンマ補正，さらにエッジ強調などを行い，最終の画質調整を行います．ただし，この段階はまだ圧縮前であり，この段階でいかに良い画質であっても，それが圧縮・伸張後も良い画質であるとは限りません．ブロック・ノイズやモスキート・ノイズといった圧縮特有のディジタル・ノイズが後から混入するためです．とくにMPEGやH.264のように圧縮率が高い場合はなおさら注意が必要です．

圧縮の前処理として重要な点は，次の4点と考えられます．

- *SN*比
- 解像度
- シャープネス
- 手ぶれ

*SN*比はそれだけでも重要な問題ですが，圧縮がからむとさらに重要になります．p.52の**図5**はノイズによって圧縮後のデータがどうなるのかを示しています．**(a)**は元画像をそのまま画面内で符号化したケース，**(b)**はその元画像にランダム・ノイズを加えたうえで，量子化係数を**(a)**と同じ値で圧縮したケース，**(c)**は**(b)**で使用した同じ画像で**(a)**と同じファイル・サイズになるように量子化係数を調整したケースの結果です．

(b)はノイズを加えたものの量子化係数は同じなので，**(a)**の画像と同じ程度のディテールを表現できています．ただし，ファイル・サイズは2倍近くまで増大してしまいました．また，**(c)**はファイル・サイズを同じにするため，量子化係数を大きくしました．そのため，ディテールは失われ，ブロック・ノイズやモスキート・ノイズが大量に発生していることがわかります．

p.53の**図6**は，720×480画素の画像に対して，解像度やシャープネスの違いが画質や生成符号量に与える影響を調べてみたものです．**(a)**は元画像をそのまま画面内符号化したケース，**(b)**はその画像にエッジ強調を加えたケース，**(c)**は**(b)**において水平方向の解像度を720から480に落としたあとで圧縮を行ったケースです．ファイル・サイズはすべて同じになるように量子化係数を調整しています．**(a)**に対して**(b)**は木の葉のディテールは良く表現されていますが，全体的にブロック・ノイズやモスキート・ノイズが増えています．**(c)**は水平方向の画素数が減っているため，ディテールの表現は**(b)**に比べて落ちていますが，全体にディジタル・ノイズが減り，見やすい画像になっているといえます．

MPEG技術者は，与えられたビット・レートの中でいか







(a) 元画像　(b) 同一量子化係数　(c) 同一ファイル・サイズ

図5　圧縮前のノイズが圧縮後のデータに与える影響

(a) は元画像をそのまま画面内符号化した．(b) はその元画像にランダム・ノイズを加えたうえで，量子化係数を (a) と同じ値で圧縮した．(c) は (b) で使用した同じ画像を (a) と同じファイル・サイズになるように量子化係数を調整した

にディジタル・ノイズを発生させずに圧縮するかということに注力していますが，この実験は圧縮する前の処理がいかにたいせつかを如実に表しているものです．圧縮する前にできる限りノイズを減らし，ビット・レートに合った解像度が得られるように周波数帯域を最適にする必要があるのです．また，それらが入力画像の状況によって適応的に変えられるのがベストでしょう．

手ぶれ補正も圧縮画像に影響を与えることがあります．先に動き補償による画質の影響について説明しましたが，手ぶれによるゆっくりとした動きがつねに存在しているため，動き予測のアルゴリズムによっては動き補償による誤差成分の発生が無視できないことがあります．これも圧縮前に低減しておけば画質は向上するでしょう．

＊　＊　＊

ここまで，MPEGのような動画圧縮画像の画質を決める要素について説明しました．圧縮する際には，

- 符号量制御
- 動き予測

の2点が重要なかぎとなるのは確かですが，圧縮の前処理でいかに画像を最適化できるかというテーマは，それと同



図6　解像度やシャープネスの違いが画質や生成符号量に与える影響
(a)は元画像をそのまま画面内符号化した．(b)はその画像にエッジ強調を加えた．(c)は(b)において水平方向の解像度を720から480に落としたあとで圧縮を行った

じか，むしろさらに重要な要素と言えるでしょう．例えばビデオ・カメラにおいて，イメージ・センサからの出力にノイズが乗っていると，圧縮によってそのノイズがさらにディジタル・ノイズを生じさせ，画質が劣化していきます．再生時にブロック・ノイズ・キャンセラやモスキート・ノイズ・キャンセラを入れ，ディジタル・ノイズを減らすことは，ある程度は可能です．しかし，同時にディテールが失われますし，再生をどのような機器で行うかわかりません．記録時にできるだけノイズを抑え，最良の状態で圧縮するべきであることに，疑問の余地はありません．

筆者らが最近発表したディジタル・カメラ/ビデオ・カメラ用の信号処理チップ「SiP-2290」は，ロー・ノイズを追求するだけでなく，フィード・フォワード制御により，画像の複雑さに適応的に対応した前処理を実行することが可能となっています．このため，MPEG-4シンプル・プロファイルというシンプルかつ軽量な圧縮でありながら，SVGA，30フレーム/sにおいて，高画質かつ3Mbpsを達成しています．

参考・引用*文献
(1) デジタルプロセッサ SiP-2290製品概要，ニューコアテクノロジー，http://www.nucore.co.jp/nu3/20_products/nucore_img_processors/digital_sip-2290.html

せい・きょうじろう
ニューコアテクノロジー(株)

<筆者プロフィール>
清 恭二郎．早稲田大学理工学部卒業後，23年間某日本民生電気機器メーカでさまざまな社内向けLSI開発に従事．しかし，その間に経験したシリコンバレーのダイナミックさにあこがれて退社．日本人の起業によるニューコアテクノロジーに参加し，アメリカン・ドリームを夢見ながらイメージ・プロセッサやMPEGのチップの開発を行っている．